

УДК 004.73; 004.77

Поцелуй А. – ст. гр. СНМ-52

*Тернопільський національний технічний університет імені Івана Пулюя*

## **ОСНОВНІ РЕСУРСИ ГЛИБИННОГО ВЕБ**

Науковий керівник: ст. викладач Маєвський О.В.

Poceluy A.

*Ternopil Ivan Pul'uy National Technical University*

## **KEY RESOURCES DEEP WEB**

Supervisor: Majevskiy A.

Ключові слова: ресурс, «глибинний веб»

Keywords: resource, "Deep web"

Проаналізуємо способи знаходження веб-ресурсів, розміщених в глибокому веб. Якщо ресурси вимагають заповнення спеціальних форм, доповнених, наприклад, капч, то необхідно вийти на базу даних, яка містить необхідні документи. Знайти бази даних – джерела прихованого веб можна за допомогою звичайних пошукових систем, узагальнивши запит і ввівши уточнюючі слова, такі як «база даних», «банк даних», «database» тощо.

Наведемо загальновідомий приклад: користувачеві потрібна статистика по катастрофам літаків в Аргентині. Природний запит до традиційної пошукової системи видає величезний список газетних заголовків. На запит «aviation database», можна відразу вийти на базу даних NTSB Aviation Accident Database.

Для пошуку в глибокому веб, а саме в тому його сегменті, який складають бази даних, сьогодні вже існують деякі спеціалізовані ресурси. Лідером серед навігаторів в глибокому веб є сайт CompletePlanet компанії BrightPlanet. Цей сайт є найбільшим каталогом, що налічує понад 100 тисяч посилань. Компанія BrightPlanet також створила персональну утиліту для пошуку в онлайн-базах даних LexiBot, яка може забезпечувати пошук в декількох тисячах пошукових систем «глибокого» веб. Метапошуковий пакет DeepQueryManager (DQM) цієї ж компанії забезпечує пошук більш ніж по 70 тисячам «прихованим» веб-ресурсам.

Дослідження, проведене ще в 2006 році [1] показало, що глибокий веб охоплює понад 300 тис. сайтів, пов'язаних з більш ніж 450 тис. базами даних, які не охоплюються традиційними пошуковими системами. До найбільш цікавих для бізнес-аналітиків ресурсів глибокого веб відносяться: бази даних юридичних і фізичних осіб; галузеві бази даних; репутаційні бази даних (чорні і білі списки); кримінологічні бази даних; бази даних товарів і послуг; каталоги продукції тощо. До всесвітньо відомих бізнес-ресурсів, розміщених в глибокому веб, відносяться: amazon.com, ebay.com, realtor.com, cars.com, imdb.com.

Наведемо приклади баз даних і каталогів глибокого веб:

FindLaw – один з найбільш популярних в світі юридичних веб-сайтів – величезний каталог правових ресурсів, що містить анотований список вільно доступних баз даних нормативно-правових документів, для яких даний ресурс є «точкою входу».

About.com – портал, що охоплює тисячі, забезпечених коментарями, посилань на веб-ресурси, в тому числі і на ресурси глибокого веб (є посилання «Invisible Web»). На порталі надається можливість пошуку в каталозі. Ресурс також включає кілька

статей з проблематики глибинного веб: «What is the Invisible Web?», «Finding the Invisible Web», «Top Places to Search the Invisible Web» та ін.

Politicalinformation.com – сервіс, що забезпечує оперативний пошук в 5000 відібраних веб-сайтів політичної спрямованості, надання новин з декількох десятків авторитетних джерел.

Infomine – сервіс забезпечує видобуток інформації з баз даних, електронних журналів (блогів), електронних доіок оголошень, електронних книг, списків розсилок, електронних каталогів тощо, переважно пізнавально-освітнього характеру. Забезпечує як загальний пошук, так і пошук по тематичним категоріям.

Особливість більшості «прихованих» ресурсів полягає у їх вузькій спеціалізації. Для пошуку в них використовуються ті ж механізми, що і для «поверхневого» веб, проте, в більшості випадків, роботи пошукових систем для глибинного веб включають унікальні для кожного такого ресурсу модулі доступу до даних.

Традиційна пошукова система найчастіше може видати адресу бази даних, але не скаже, які документи конкретно містяться в ній. Типовий приклад – інформаційно-пошукові системи з українського (zakon.rada.gov.ua) законодавства. Тисячі документів з баз даних стають доступні тільки після входу в систему, а роботи стандартних пошукових систем не в змозі заіндексувати контент баз даних.

Парадоксально, але в якості одного з ресурсів глибинного веб можна розглядати і архів матеріалів відкритого веб-простору. Такий архів – Internet Archive з 1996 року створює компанія Alexa. Сьогодні обсяг бази даних Alexa перевищує 350 млрд. веб-сторінок.

Технологія сховища Alexa включає ряд сучасних засобів управління гігантським документальним сховищем. Наприклад, за допомогою технології Alexa виконується кластеризація веб-ресурсів, тобто формування колекцій документів, близьких по тематиках. Особливий інтерес у користувачів сервісу Alexa викликає «Машина часу» (Wayback Machine), що відкриває доступ до тимчасових зрізів веб-простору. Одне з найбільш цікавих практичних застосувань цієї технології – відновлення документів, колись опублікованих в веб-просторі, але згодом видалених. При цьому зростання глибинного веб загрожує серйозними проблемами повноти в сховищі системи, пов'язаними зі зростаючою кількістю сайтів, які експлуатують різні типи технологій управління контентом, динамічною публікацією документів з баз даних тощо.

Література.

1. He B., Patel M., Zhang Z., Chang K. C.-C. Accessing the Deep Web: A Survey // Communications of the ACM (CACM), 50(5):94-101, 2007.

УДК 621.326

Рокош М.– ст. гр. СНм-51

*Тернопільський національний технічний університет імені Івана Пулюя*

## **ДОЦІЛЬНІСТЬ ВИКОРИСТАННЯ NOSQL**

Науковий керівник: к.т.н., доцент Загородна Н.В.

M. Rokosh

*Ternopil Ivan Pul'uj National Technical University*

## **THE REASONABILITY OF USING NOSQL**